DOCUMENT RESUME

ED 232 835 RC 014 253

AUTHOR Gillespie, Francis P.; And Others
TITLE Public Policy and Ethnic Codes: A Sensitivity Analysis of Hispanic Vital Statistics.
INSTITUTION Texas Univ., Austin. Population Research Center.
SPONS AGENCY Ford Foundation, New York, N.Y.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.
PUB DATE Oct 82
GRANT 805-0799; HD-06160-11A1
NOTE 16p.; Paper presented at the Annual Meeting of the Southern Regional Demographic Group (13th, Greensboro, NC, October 6-8, 1982).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Catholics; *Census Figures; Church Workers; *Data Collection; *Ethnic Origins; Females; *Hispanic Americans; Identification; *Labeling (of Persons); Mexican Americans; Research Methodology; Statistical Bias; Statistical Data
IDENTIFIERS *Death Records; *Texas

ABSTRACT

Data from death records of 1,906 Catholic parishes and other entities in 11 Texas dioceses during January to June 1980 were used to estimate sensitivity of Hispanic population data to various identifications of "Hispanic." Self-identification as Hispanic, as on the 1980 census, appeared unreliable; the Census Bureau was expected to recode 1980 census records according to an expanded list of Spanish surnames. The new Texas death certificate asked for father's surname and "Was the decedent of Spanish origin?", but less was known about third-party identification in vital statistics than about self-identification. To check whether Census Bureau and Texas data for estimating Hispanic population accurately identified persons whom local sources classified as Hispanic, questionnaires, accompanied by letters from local bishops encouraging cooperation, were mailed to Texas Catholic pastors, who knew the decedents' backgrounds, asking them to identify Hispanic decedents. Response rate was 100%; 8,163 deaths were tabulated and compared statistically to data from the Texas Bureau of Statistics GUESS computer program, which used linguistic principles to determine origin of surname, and the Census Bureau's improved list of Spanish surnames. Census Bureau and Texas classification programs yielded remarkably similar results; local third-party definition yielded a larger number of Hispanics, especially Hispanic women. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED232835

AUG 1983
RECEIVED
ERIC/CRESS

PUBLIC POLICY AND ETHNIC CODES: A SENSITIVITY ANALYSIS OF HISPANIC VITAL STATISTICS

Paper prepared for the 13th annual
meeting of the Southern Regional Demographic Group
Greensboro, N.C., October 6-8, 1982

Francis P. Gillespie, S.J.
St. Joseph's University, Philadelphia
Andrew M. Greeley
University of Arizona, Tucson, and NORC
Michael Hout
University of Arizona, Tucson
Teresa A. Sullivan
University of Texas, Austin

The names of the authors are in alphabetical order. The authors wish to acknowledge the programming assistance of Shawn Boyd and the assistance of Kathleen Elliott, who typed this paper. This research was funded by Ford Foundation grant 805-0799 and by NICHD Grant HD 06160-11A1 to the Population Research Center at The University of Texas at Austin.

RC014258

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Teresa A. Sullivan
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Vital Statistics as a Special Problem

Demographic data classified by race or ethnicity are mandated by federal and state statutes for many programmatic uses. Using grouped racial/ethnic data is usually considered justifiable in the current political climate. However, official identification of individuals as members of specific racial/ethnic groups continues to be suspect. The group-versus-individual tension has been resolved through a two-pronged approach: first, voluntary of self-identification approaches to racial/ethnic categories have largely replaced earlier approaches; secondly, the data have been kept confidential so as to minimize the risk of associating individual names with racial/ethnic codes.

But these solutions have their disadvantages. The self-identification system currently used by federal statistical agencies is criticized by one statistician for its "extraordinary ambiguity" (Kruskal, 1981: 512; see also Lowry, 1980). Yet the same writer realizes the political and legal pitfalls of administrative determination of race or ethnicity (Kruskal, 1981: 513, n.5). Furthermore, confidentiality works imperfectly with vital statistics data. Census and survey data may be kept confidential because no third party may legitimately associate the individual's name with his (or her) data record. Vital statistics are different in important ways. The birth certificate legally defines the native-born citizen and may be legitimately required to produce evidence of age or citizenship. Similarly, the death certificate is needed to verify the characteristics of the decedent so that heirs may collect insurance, or other benefits, transfer the estate's property, and so on. Although demographers use vital statistics data only in the aggregate, any stigmatizing information contained there could harm individuals.

A recent Louisiana lawsuit highlights the potential conflict between individual rights and collective data needs. In this lawsuit, a Louisiana woman contested the state's designation of a child as "black" on the birth certificate if the child were at least one-eighth black. The case raised the spectre of the old black codes of the South, which defined "blackness" relative to ancestry, and of the more recent horrors of the Nuremberg laws. Regardless of the need for collective data on racial differentials in vital rates, the plaintiff protested her personal, permanent identification as a member of a racial minority.

Unlike Louisiana, most states (as well as Federal statistical agencies) make little effort to define racial or ethnic categories. The system commonly used is called "self-identification," but the term is misleading, especially with vital statistics. Race or ethnicity is "self-identified" when respondents answer for themselves, as in a census or survey. But in many cases, respondents answer for other members of their households. In the case of vital records, some third party must answer for the newborn child or for the decedent. In an unknown percentage of cases, the physician, medical examiner, funeral director, or other informant supplies the information. How often this third party consults the next of kin is unknown.

Thus vital statistics data pose particular problems for both preserving the confidentiality of race/ethnic data and for developing valid and reliable classifications.

Classifying Hispanics as a special issue

If race/ethnic codes in general pose a difficult problem, the problem is intensified for members of Hispanic ethnic groups. Since 1930, the U.S. Census Bureau has used a bewildering variety of variables for identifying Hispanics. Some of them include race, self-identified ancestry, language

spoken, place of birth, parental nativity, and surname. (For a review, see Hernandez, Estrada, and Alvirez, 1973). The overlap among these classification schemes is substantial in some parts of the country, but is by no means uniform (Siegel and Passel, 1979). Moreover, there is indication that during the decade from 1970 to 1980, the process of self-identifying oneself as Hispanic changed, so that the self-identification variable does not appear to be reliable. (U.S. Census, 1974: Siegel and Passel, 1979:9).

With the Hispanic population growing rapidly through both immigration and natural increase, population estimates and projections require estimates of vital rates, and that in turn requires Hispanic identification in vital statistics. But until quite recently, only a few states coded Hispanic ethnicity at all, and those that did used unique coding systems. (Sullivan Gillespie, Greeley, Hout, 1981: 27-28).

Since 1980, the National Center for Health Statistics has encouraged the states to include an Hispanic identifier on vital records. The new Texas death certificate (appendix A) contains an example of the identifier that many states added. Although it is superficially similar to the self-identification question used in the 1980 Census, it is answered by third parties. We know even less about how the third-party identification works in vital statistics than we do about self-identification in census data. Moreover, because of the personal identification aspect of vital records mentioned earlier. it is not yet clear whether this identifer will be as acceptable as that on the census questionnaire. At least one influential Hispanic citizens' group has protested the identifier on the Texas birth and death certificates.

We are not able to determine the reliability and validity of the new "self-identified" Hispanic datum on birth and death certificates. What we are able to do, because of a unique data set, is to estimate the sensitivity of Hispanic population data to various definitions of "Hispanic." Our data

set contains a third-person Hispanic identifier that might provide a proxy for the one in actual use.

Data and Methods

The data reported here are from the death records of the Catholic Church in Texas during the period January 1 - June 30, 1980. (The U.S. Census fell mid-way between those two dates.) A book of death records is kept in every Catholic parish (or other entity) that conducts funerals; it contains the name of the decedent, age at death, name and address of nearest relative, date of death and of burial, the attending priest, sacraments received near death (e.g., the final anointing), place of burial, and remarks. We collected our data by mailing pastors a questionnaire asking for specific information or decedents from the death records book. Each questionnaire was accompanied by a cover letter from the local bishop encouraging cooperation. At the time of the study, there were 11 dioceses in Texas, each of which gave us permission to conduct the study. Over 1900 entities were surveyed:

897	parishes
330	Missions
73	stations
509	chapels
30	hospitals
38	motherhouses, provincialates, or other headquarters of religious groups
29	other institutions (e.g., nursing homes, orphanages)
1906	entities

The response rate was 100%, and 8163 deaths were tabulated.

Of all Hispanic groups in the country, Mexican-Americans are the most likely to report themselves as Catholic. Of the Mexican-origin respondents to the 1971-1977 General Social Surveys, 93.6% were reared as Catholics and another 1.2% converted to Catholicism. The General Social Surveys are representative only of the English-speaking population over the age of 18. One might expect that among Spanish-speaking Mexican-Americans, and those still

in the parental household, the percentage Catholic might be even higher. Hispanics in Texas are overwhelmingly of Mexican origin. Based simply on these data, it seems likely that Texas parish records would cover much of the Hispanic population.

However, many Hispanics are not regular church attenders; the General Social Surveys show that 5.8% of Mexican Americans report that although they were reared Catholics, they now profess no religion.

Between 7 and 10 percent of Southwestern Hispanics are believed to be non-Catholics. While some of these persons are probably buried in the Church by Catholic relatives, many are not. Moreover, some aged or ill Mexican-American may return to Mexico to die. A demographic analysis of annualized parish data, compared with Texas vital statistics data, suggests that our data set covered about 76% of the Hispanic deaths as classified by the Texas Bureau of Vital Statistics. (Gillespie, Greeley, Hout and Sullivan, 1982). While this does not suggest that our data are an adequate substitute for death certificate data, it does suggest that our data represent a large fraction of the target population, and a comparison of identifiers for this target population is a useful exercise.

We used three techniques to identify Hispanics. The first was the "local" technique, which was the pastor's identification of the decedent as Hispanic. Pastors and assistant pastors had recently officiated at the wakes, funerals, and burial services of the decedents. They had talked with the family and met the other mourners. In many cases, they had known the decedents for years, and had visited many on their deathbeds. Thus, the pastor is in an excellent position to know the background of the decedent.

This may be illustrated by an anecdote. One decedent named Pinala was not listed as Hispanic on a questionnaire we received, even though he lived in a predominantly Hispanic area and Pinala is a common Spanish surname.

"Why did you indicate that Mr. Pinala was not Hispanic?" our investigator asked the pastor during a routine editing call-back. "Because he was born in Italy," said the pastor. "He spoke Italian, he ate Italian food, his family is Italian, and he thought he was Italian. That's why I didn't call him Hispanic."

The second identity code, the "state" code, is based on a computer program written by Dr. Robert Buechley and called "GUESS." GUESS uses linguistic principles to determine the origin of surname; it is used by the Texas Bureau of Vital Statistics.

The third identifier, the "national" identifier, is based upon the recently expanded and improved list of Spanish surnames used by the U.S. Census Bureau (Passel and Word, 1980). The Census Bureau is expected to recode 1980 census records according to Spanish surnames in five Southwestern states (Arizona, California, Colorado, New Mexico and Texas) (Siegel and Passel, 1979:41).

Results

As Table 1 shows, for a large sample of deaths there was 93.2% agreement among the three techniques for male decedents, and 91.1% agreement for female decedents. If we exclude the approximately one-third of deaths that were agreed by all techniques to be non-Hispanic, then there was 92.4% agreement that the remaining males were Hispanic, and 88.1% for females.

An analysis of the three combinations of two techniques provides indicators of the sensitivity of current estimates. In Texas, the current estimates are based on surname analysis using the Buechley program, GUESS. The ratio of deaths classified Hispanic by the census list of surnames, to deaths classified Hispanic using GUESS, is used to adjust the number of deaths the state reports as Hispanic. The adjusting ratios are 0.97 for males and 0.98 for females, an indicator that the national Spanish surname list produces a

smaller estimate. This estimate is called the "national" estimate in Table 2.

The ratio of deaths classified Hispanic by the local informant to deaths classified Hispanic by the state program, is the second adjustment ratio. This "local ratio" is 1.02 for males and 1.05 for females. A value greater than one indicates that pastors are more likely than the GUESS program to classify someone as Hispanic, probably because the pastor has much more information than surname on which to base the estimate. The higher estimate for females relative to males suggests that some married women may no longer bear recognizable Spanish surname. While this comparison does not tell us how the Hispanic identifier on the death certificate will work, it is one indicator that informed third parties might give somewhat higher estimates than the state surname identifier.

The final comparison given in Table 2, "union," is the ratio of deaths classified Hispanic by any of the three techniques, to the deaths classified Hispanic by the state program. As might be expected, this yields the largest ratios, 1.04 for males and 1.08 for females. The range in deaths to assumed Hispanics is approximately 495 for both males and females, but the female range amounts to 9.8% of the total female Hispanic deaths recorded by the state in 1980.

[The remaining possible combination of techniques, the ratio of the deaths classified Hispanic by the local respondent, to the ratio of deaths classified as Hispanic by the census, yields ratios of 1.05 for males and 1.08 for females. While these ratios could hypothetically be used to adjust the 1980 Spanish-surname census data for Texas, those data are not available at this writing. Furthermore, this procedure would require an additional assumption that the distribution of Spanish surnames among decedents is similar to that in the enumerated population. For a variety of reasons,

especially the age structure of decedents and the more recent trend toward Hispanic exogamy, this assumption does not appear to be justified.]

Conclusion

The recoding of Spanish surnames offers a compromise position between administrative definition of ethnic group and an ambiguous self-identification. Because the Spanish-surname recode is used only in statistical tabulations and does not appear on the vital record itself, its possible stigmatizing influence is neutralized. The two classification programs used by the federal government and by the state of Texas yield remarkably similar results; a local third-party definition, by contrast, yields a larger number of Hispanics, especially Hispanic women.

Siegel and Passel (1979) argue that the Hispanic surname is a reasonably good identifier in the southwest and suggest its use in other administrative records. This study leads to a similar conclusion, with the caveat that the surname identifier needs to be monitored carefully for women. If women disproportionately give up their Spanish surname upon marriage, and if exogamy is a cohort phenomenon, then the deaths for Hispanic women will be undercounted and their death rates will be too low. One alternative worth further consideration is the coding of the father's surname for married women.

TABLE 1. Cross-classification of parish records by local, state, and national identifiers of Hispanic background, four Texas dioceses, 1980.

Is decedent Hispanic?			Percent of records	
Local	State	National	Male	Female
Yes	Yes	Yes	60.6	54.1
No	Yes	Yes	0.3	0.6
Yes	No	Yes	0.4	0.7
No	No	Yes	0.3	0.4
Yes	Yes	No	1.9	2.0
No	Yes	No	0.4	0.4
Yes	No	No	1.6	3.2
No	No	No	32.6	37.0
Unknown			1.8	1.6
		Δ=6.8	99.9%	100.0%

Table 2

Sensitivity of Texas data using changes in Hispanic definition, by Sex, 1980.

Definition[a]	Numbers of deaths	Crude death rate[b]
Males		
National	7612	5.1
State[c]	7810	5.2
Local	7971	5.4
Union	8107	5.4
Range of estimates	495	0.3
	(6.3% of State)	
Females		
National	4914	3.3
State[c]	5029	3.4
Local	5284	3.5
Union	5408	3.6
Range of estimates	494	0.3
	(9.8% of State)	

SOURCE: Texas Department of Health, Bureau of Vital Statistics, Unpublished data; U.S. Bureau of the Census, Unpublished data; Texas Catholic Parish record study.

NOTES: [a]See text for definition.

[b]Rate per 1,000. Denominator used was Spanish-origin population as of the 1980 Census date (self-identified).

[c]The number reported by the Texas Bureau of Vital Statistics.

APPENDIX A

BEST COPY AVAILABLE

STATE OF TEXAS

CERTIFICATE OF DEATH

STATE FILE NO

Texas Department of Health — BUREAU OF VITAL STATISTICS

1 NAME OF DECEASED [Type or print] [a] First [b] Middle [c] Last | 2 SEX | 3 DATE OF DEATH

4 RACE | 5a WAS THE DECEDENT OF SPANISH ORIGIN? | 5b IF YES SPECIFY MEXICAN, CUBAN, PUERTO RICAN, ETC. | 6 DATE OF BIRTH | 7 AGE [In years last birthday] | IF UNDER 1 YEAR: Months, Days | IF UNDER 24 HRS.: Hours, Minutes

8a PLACE OF DEATH — COUNTY | 8b CITY OR TOWN [If outside city limits, give rural] | 8c NAME OF [If not in hospital, give street address] HOSPITAL OR INSTITUTION | 8d INSIDE CITY LIMITS?

9 MARRIED, NEVER MARRIED, WIDOWED, DIVORCED [Specify] | 10 BIRTHPLACE [State or foreign country] | 11 CITIZEN OF WHAT COUNTRY? | 12 WAS DECEDENT EVER IN U.S. ARMED FORCES? | 13 SURVIVING SPOUSE [If wife, give maiden name]

14 SOCIAL SECURITY NO | 15a USUAL OCCUPATION [Give kind of work done during most of working life, even if retired] | 15b KIND OF BUSINESS OR INDUSTRY

16a RESIDENCE — STATE | 16b COUNTY | 16c CITY OR TOWN [If outside city limits, show rural] | 16d STREET ADDRESS [If rural, give location] | 16e INSIDE CITY LIMITS?

17 FATHER'S NAME | 18 MOTHER'S MAIDEN NAME | 19 SIGNATURE OF INFORMANT

CAUSE OF DEATH

PART I — Conditions, if any, which gave rise to immediate cause stating the underlying cause last

(a) IMMEDIATE CAUSE [Enter only one cause per line for (a), (b), (c)] | Interval between onset and death

(b) DUE TO, OR AS A CONSEQUENCE OF | Interval between onset and death

(c) DUE TO, OR AS A CONSEQUENCE OF | Interval between onset and death

PART II — OTHER SIGNIFICANT CONDITIONS — CONDITIONS CONTRIBUTING TO DEATH BUT NOT RELATED TO CAUSE GIVEN IN PART I (a) | 21. AUTOPSY?

22a ACC., SUICIDE, HOM., UNDET., OR PENDING INVEST. [Specify] | 22b DATE OF INJURY [Mo., Day, Yr.] | 22c HOUR OF INJURY M | 22d DESCRIBE HOW INJURY OCCURRED

22e INJURY AT WORK [Specify yes or no] | 22f PLACE OF INJURY — At home, farm, street, factory, office building, etc. [Specify] | 22g LOCATION — STREET OR R.F.D. NO. — CITY OR TOWN — STATE

CERTIFIER

To be completed by CERTIFYING PHYSICIAN only

23a To the best of my knowledge, death occurred at the time, date, and place and due to the cause(s) stated [Signature and Title]

23b DATE SIGNED [Mo., Day, Yr.] | 23c HOUR OF DEATH M

23d NAME OF ATTENDING PHYSICIAN [Type or print]

To be completed by MEDICAL EXAMINER or CORONER only

24a On the basis of examination and/or investigation, in my opinion death occurred at the time, date, and place and due to the cause(s) stated [Signature and Title]

24b DATE SIGNED [Mo., Day, Yr.] | 24c HOUR OF DEATH M

24d PRONOUNCED DEAD [Mo., Day, Year] ON | 24e PRONOUNCED DEAD [Hour] AT M

25a BURIAL, CREMATION, REMOVAL [Specify] | 25b DATE | 25c NAME OF CEMETERY OR CREMATORY

25d LOCATION [City, town, or county] [State] | 26 SIGNATURE OF FUNERAL DIRECTOR OR PERSON ACTING AS SUCH

27a REGISTRAR'S FILE NO | 27b DATE REC'D BY LOCAL REGISTRAR | 27c SIGNATURE OF LOCAL REGISTRAR

VS-112, REV. 1/80

SAMPLE

13

14

REFERENCES

Gillespie, Francis P., Andrew M. Greeley, Michael Hout, and Teresa A. Sullivan.

1982 "Texas Catholic Parish Records as Auxiliary Vital Statistics". Paper presented at the 1982 annual meeting, the Southwestern Social Science Association.

Hernandez, Jose, Leo Estrada, and David Alvirez.

1973 "Census data and the problem of conceptually defining the Mexican American population." Social Science Quarterly 53 (March): 671-87.

Kruskal, William

1981 "Statistics in society: problems unsolved and unformulated." Journal of the American Statistical Association, 76 (Sept. 1981): 505-515.

Lowery, Ira S.

1980 "The Science and politics of ethnic enumeration," Santa Monica, California: Rank Paper Series, No. P-6435.

Passel, Jeffrey and David L. Word

1980 "Constructing the list of Spanish surnames for the 1980 census: an application of Bayes' theorem." Paper presented at the 1980 Annual Meetings, Population Association of America.

Siegel, Jacob S. and Jeffrey Passel

1979 "Coverage of the Hispanic population of the United States in the 1970 census." Current Population Report, Series P-23, No. 82, Washington: USGPO.

Sullivan, Teresa A., Francis P. Gillespie, Andrew M. Greeley, and Michael Hout

1981 Estimating Hispanic Undercount in the 1980 Census Through the Use of Catholic Parish Records. Report to the Ford Foundation.

U.S. Bureau of the Census

1974 "Consistency of reporting of ethnic origin in the Current Population Survey." Technical Paper n. 31.